

# ***CY15 Livermore Computing Focus Areas***

*Tom Connell, Kim Cupps, Trent D'Hooge, Tim Fahey,  
Dave Fox, Scott Futral, Mark Gary, Robin Goldstone,  
Pam Hamilton, Todd Heer, Jeff Long, Rich Mark,  
Chris Morrone, Jerry Shoopman, Joe Slavec, David Smith,  
Becky Springmeyer, Marc Stearman, Py Watson*

LLNL-TR-666561

Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344.



## **Disclaimer**

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

## ***Contents***

<b><i>Purpose.....</i></b>	<b><i>1</i></b>
<b><i>Executive Summary .....</i></b>	<b><i>1</i></b>
<b><i>Process.....</i></b>	<b><i>1</i></b>
<b><i>CY15 Drivers.....</i></b>	<b><i>1</i></b>
<b><i>LC CY15 Focus Areas.....</i></b>	<b><i>2</i></b>
<b>Improving Service Quality .....</b>	<b>3</b>
Non-Purgeable Disk Space.....	3
Time to First Archive Byte .....	3
Tool Development and Improvement.....	3
Disaster Recovery .....	5
Lustre File System Improvement and Hardening .....	5
Virtualization Environment Improvements .....	5
<b>Monitoring, Automation, Delegation and Center Efficiency.....</b>	<b>6</b>
Monitoring.....	6
Automation .....	7
Delegation and Efficiency.....	8
<b>Next Generation Compute and Data Environments.....</b>	<b>10</b>
CTS-1 Preparation .....	10
Sierra Preparation .....	10
Pre-Exascale Preparation.....	11
Big Data .....	11

## Purpose

This document identifies and describes three cross-cutting areas demanding the particular attention and focus of Livermore Computing (LC) during Calendar Year 2015 (CY15). Identifying these focus areas allows LC management and staff to make sound decisions about the application of effort and expenditures throughout the year. ***This document is expressly not intended to be a comprehensive description of LC efforts.*** Rather, it shines a light on high priority issues involving multiple LC disciplines.

## Executive Summary

The LC team undertook a survey of primary Center drivers for CY15. Identified key drivers included enhancing user experience and productivity, pre-exascale platform preparation, process improvement, data-centric computing paradigms and business expansion. The team organized critical supporting efforts into three cross-cutting focus areas:

- Improving Service Quality
- Monitoring, Automation, Delegation and Center Efficiency
- Next Generation Compute and Data Environments

In each area the team detailed high level challenges and identified discrete actions to address these issues during the calendar year. Identifying the Center's primary drivers, issues, and plans is intended to serve as a lens focusing LC personnel, resources, and priorities throughout the year.

## Process

This document is the result of a two month series of weekly meetings involving representatives from throughout LC. These disciplines included: networking, visualization, security, customer services, development environment, archive storage, systems software, operations, the Advanced Technology Office (ATO), and file systems. The meetings explored each discipline area in turn, focusing on issues and efforts as they related to primary LC drivers. Each area presented to the assembled team allowing for an extensive interchange of information both amongst and across disciplines. The team then identified critical areas requiring focused attention during CY15.

## CY15 Drivers

The LC team identified the following primary drivers for CY15:

**User Experience and Productivity:** Enhancing the productivity of LC users and providing a positive and productive experience remains a cornerstone of LC's mission. This driver includes providing high availability and reliability in support of LC customers.

**Pre-Exascale Machine Preparation:** The next five years will witness the procurement and deployment of two new machine architectures within LC; namely the Commodity Technology System (CTS-1) machines and the Sierra Advanced Technology System (ATS-2). The Sierra system will strain or break many of today's High Performance Computing (HPC) software products and technologies, and will compel LC to make significant changes to the solutions fielded today. Making these machines productive for LC customers requires significant upfront investment and preparation throughout CY15. In addition, teams from LLNL and DOE Laboratories across the country are engaged in advanced technology research and development efforts in preparation to support machine architectures and environments beyond the Sierra machine timeframe.

**Process Improvement - Monitoring, Automation, Delegation and Efficiency:** The number and variety of environments and systems supported by LC has grown markedly in the last five years (Collaboration Zone, Restricted Zone, Infrastructure Zone, iHPC, SNSI, Global Security, Green Collaboration Environment, etc.). During this same time both staffing and funding has dropped significantly. Every LC discipline will need to identify ways to improve, streamline and automate internal processes in order to accommodate existing environments and to facilitate growth into new mission areas. Efficiencies in the areas of automation and monitoring are required as is the enablement of first level responders (Hotline and Operations) to independently respond to and bring to resolution commonly encountered issues and tasks.

**Data-centric Computing Paradigms:** The location and use of Non-Volatile RAM (NVRAM) in upcoming compute architectures and the prospect of mission expansion in the area of Big Data in support of science shine a light on the need for LC to grow its knowledge-base in this area. The arrival of the Catalyst machine has provided a platform for LC staff, CASC researchers, and LC customers, who can now delve into data management and operational models made possible by these machines. The support of data-driven event detection, correction and correlation is presenting LC with its own internal big data problems to solve.

**Best Practice Measurement and Synthesis Effort:** LC suffers today from not being able to adequately track improvements or degradations after deployments or configuration changes in our environments. This impacts decision-making and may lead to poor investments in hardware, software, and manpower.

**Business Expansion (Work For Others - HPC-IC, CRADA, SNSI, Global Security):** Growing Center business is key to LC's future. We must find ways of leveraging the significant investments that have been made in the SNSI environment, the HPC-IC, and the significant preparation for a Cooperative Research And Development Agreement (CRADA) in power management, file system, STAT and Spindle software product areas.

## LC CY15 Focus Areas

Three cross-cutting (multidisciplinary) focus areas were called out for the primary attention of LC staff during the coming year. These areas were:

- Improving Service Quality
- Monitoring, Automation, Delegation and Center Efficiency
- Next Generation Compute and Data Environments

Below, each focus area is briefly introduced, followed by a description of the area's key contributing issues and finally, a description of LC's planned approach to address the issues during the year.

## **Improving Service Quality**

Supporting high levels of user productivity is at the core of LC's mission. Providing consistent, quality services across a growing number of environments and mission spaces requires development of improved efficiencies and the elimination of wasted effort. A number of areas for improvement have been identified that will enhance our customers' and our own experience and productivity while providing a high return on investment.

### ***Non-Purgeable Disk Space***

#### *Issue:*

Users frequently request expansion of NFS file system capacity and quotas so that they can store data that is not subject to scratch file system purge policies. In the past, extra NFS space has been allocated to select users as "project space" to satisfy pressing needs. Unfortunately significant expansion of existing (backed-up) NFS file systems is cost-prohibitive at the scale required by LC users.

#### *Approach:*

During CY15 LC will undertake a study to determine the proper file system software and hardware platforms to best support a significant amount of non-purgeable space for users. The chosen technologies will be procured and deployed and allocation policies will be established and advertised to LC users.

### ***Time to First Archive Byte***

#### *Issue:*

Archive users often experience excessive delays accessing files in HPSS archives even when the target files have been recently written. As a result, some customers are not leveraging archive capabilities claiming that it is cost/time-prohibitive to retrieve data from tape. While LC's HPSS archives provide disk cache front-ends with excellent time to first byte performance, these disk caches haven't grown in years, and are too small to store files online for a significant amount of time.

#### *Approach:*

In CY15 HPSS disk caches will be doubled in capacity. LC studied historical and CY15 archive growth projections, and forecast that CY15 disk cache expansions will enable files written to the archives to remain resident on disk for many (6-12) months, regardless of file size. The expanded disk caches will also enhance the robustness and reliability inherent with the HPSS archives and will allow disk data to be migrated to tape optimally, requiring fewer archive tape drives.

### ***Tool Development and Improvement***

#### *Issue:*

LC customers leverage a wide range of LC-developed tools to optimize their daily work and production work-flows. LC has identified a number of high Return On Investment (ROI) improvements to LC-authored tools and capabilities that will further enhance user productivity.

*Approach:*

High ROI tool investments for CY15 include:

- The highly popular Hopper data management tool performs file transfers, searches, directory synchronization, disk usage exams, and other tasks via a simple drag-and-drop interface. In CY15 Hopper will be enhanced to include “turbo” copy and sync operations which will improve throughput on large-scale copy and synchronization operations. In addition Hopper will integrate support for the locally developed parallel file system tools dcp, dwalk, and drm.
- The Spack package management tool supports multiple versions and configurations of software on a wide variety of platforms and environments. During CY15 Spack’s ability to handle multi-laboratory and multi-operating system environments will be greatly enhanced to support BGQ and possibly Cray. Also to be added will be optional dependencies, Python package support, build and test options for packages, and improved handling of LLNL MPI builds.
- Spindle is a tool for improving the library-loading performance of dynamically-linked applications. In particular Spindle provides scalable loading of shared libraries, executables and python files at scale without turning the file system into a bottleneck. In CY15 Spindle will be ported to the BGQ architecture. Further development of Spindle will be supported in part by CRADA funding from Intel Corporation. In addition the Spindle tool will be supported for and promoted to customer application teams.
- Lorenz makes HPC easier for LC users through the use of modern web technologies. The MyLC dashboard is a large part of this effort, and is a customizable view (via portlets) into the Center from a user’s perspective. During CY15 Lorenz developers will implement a repository that tracks Center cluster changes and a portlet allowing users to display this history. In addition, a subscription service will be developed allowing users to subscribe to notification of particular events. Additional work will extend existing portlets to display more information to users about file systems and their usage.
- Performing efficient file system operations (e.g., copy and remove) across large numbers of files and directories is time consuming and often very inefficient with existing tools. LC has developed and is testing four tools which maximize performance of these operations through parallelism. These tools are:
  - dwalk - recursive walk of a directory and display information
  - dcp - recursive parallel copy
  - dcmp - recursive parallel compare
  - drm - recursive parallel remove

In CY15 the testing of these tools will be completed and they will be placed into production.

## ***Disaster Recovery***

### *Issue:*

In recent years there has been an emphasis on site-wide and Tri-Laboratory Disaster Recovery. Less emphasis has been placed on localized disaster recovery scenarios such as single building loss or the loss of individual high-value infrastructure components.

### *Approach:*

In CY15 LC will perform an audit of infrastructure Disaster Recovery vulnerabilities. Vulnerabilities will be listed and prioritized and mitigations put in place as appropriate.

## ***Lustre File System Improvement and Hardening***

### *Issue:*

The largest challenge, and probably the most well-known pain point for LC users, is the availability, reliability, and performance of LC's Lustre parallel file systems. These challenges directly impact the ability of LC customers to do work.

### *Approach:*

There are limitations as to how much impact LC can have on Lustre stability and performance. In the area of stability we will enhance the local Lustre software test suite and lead the global community effort to improve Lustre development practices, enhance testing, and documentation of Lustre protocols. LC I/O experts will employ the Darshan HPC I/O characterization tool, targeting high-profile and suspect applications in order to understand and improve application access patterns to Lustre.

MetaData Server (MDS) performance is at the center of the majority of Lustre performance complaints. During CY15, LC Lustre developers, supported by a CRADA with Intel Corporation, will work on one area that promises the highest return on investment, namely large dnode support. Large dnode support will expand the size of an internal data structure obviating the need for multiple accesses for commonly accessed data.

## ***Virtualization Environment Improvements***

### *Issue:*

Over the last two years, LC consolidated many of its unclassified infrastructure servers onto a unified hardware (Cisco UCS 5108) platform via server/service virtualization. This process has been at times difficult and we have had to refactor our environment multiple times. We have also found the need to stand up a VMware environment alongside our UCS systems (currently running RHEV) in order to support our RSA two-factor authentication service. Much of our virtualization expertise resides in just one or two individuals and our deployment experiences have taught us a number of lessons and revealed areas for improvement in the creation and deployment of new Virtual Machines (VM) environments.

### *Approach:*

In CY15 LC will undertake a comparison of KVM/RHEV and VMware Virtual Machine (VM) platforms using the knowledge we've gained over the last two years. This study will help inform our path forward in the area of virtualization. We will also address the issue of deprecated support for RHEV surrounding TOSS software rollouts.

Before LC can consider offering VMs in support of end-user LC customer requirements, it must first enhance and spread throughout LC its virtualization expertise. In CY15 LC will expand the breadth of its VM expertise and cross-education across system administration staff. Using this education and lessons learned we will undertake an effort to streamline VM creation and deployment.

## **Monitoring, Automation, Delegation and Center Efficiency**

The LC environment is complex and is comprised of over two hundred different compute and infrastructure systems. Monitoring the health and performance of these systems and their interoperability is extremely challenging. Providing proper automated response to issues detected via monitoring is even more difficult. In the face of increasing complexity and decreasing or flat staffing levels, LC finds itself searching for new solutions capable of rapidly detecting failures and performance regressions and taking appropriate actions in an automated fashion. LC is also searching for efficiency through empowering first responders (LC Hotline and Operations) with the ability to detect and correct more commonly encountered issues without the need for involvement of other Center staff (termed “delegation” here).

### ***Monitoring***

#### *Issue:*

LC utilizes a diverse set of monitoring and logging tools to track the health and status of its HPC systems and infrastructure. Skumme is the “single pane of glass” which presents the best overall picture of Center health. Over the last year, Skumme integration with other tools, including the Splunk log processing tool, has dramatically improved our ability to monitor certain classes of Center issues. However, there exists significant room for improvement in the areas of problem discovery and providing real-time monitoring alerts.

#### *Approach:*

At the heart of monitoring lies real-time data collection and parsing. In CY15 LC will enhance its collection of Splunk queries and automated alerts and actions. As new operational issues are encountered where automated flagging is appropriate, administrators will enhance and adapt Splunk-driven monitoring capabilities. Splunk will be integrated into production in SNSI and Green Collaboration environments and system administrators across LC will be encouraged to grow LC’s toolset of automated alerts.

The Lightweight Distributed Metric Service (LDMS) provides a mechanism for extremely lightweight and scalable collection of run-time data. LDMS is distributed with current TOSS operating system releases and has the potential for providing a wealth of information that can be parsed by monitoring applications. During CY15 LC will investigate integrating selected LDMS data streams with LC monitoring applications.

In LC’s InfiniBand environment, the OpenSM is used for both subnet management and subnet administration. As InfiniBand issues often heavily impact Center operation, LC will expand the OpenSM monitoring service and integrate those services with Skumme during CY15.

Monitoring and advertising service uptime metrics, platform utilization, and Center-wide statistics is important for LC Staff and Center management's insight into Center-wide health. Synthesis of this information often identifies opportunities to improve LC services. In CY15 Lustre uptime metrics will be gathered and published for the consumption of Center staff and management. Lastly, the gathering of a subset of the current Excel-based monthly archive statistics will be automated, and results will be displayed in a MyLC portlet to enhance graphics and enable trending.

## **Automation**

### *Issue:*

LC staff is often stretched by the scale and complexity of the numerous LC environments. Automating tasks and responses to normal and off-normal conditions is required in order to maintain and enhance service quality. While a myriad of LC tasks are already automated, there are many areas that require attention.

### *Approach:*

In CY15 LC will enhance Center automation via the following efforts:

- Compliance Tool Generation - During new software rollouts, and as part of regularly scheduled mandated checks, system administrators spend an inordinate amount of time performing security compliance checking (e.g. manual SSCL tests). During CY15 a student will be tasked with the automation of many of these tasks through the development of a compliance tool.
- Semi-Automated System Shutdown - During cooling emergencies it is vital that various ranges of equipment be powered down quickly. Currently, operators and administrators log in to each Cluster individually and issue commands to shut down hardware; this is very time-consuming. In CY15 a suite of shutdown scripts will be developed and integrated into Skumme so that an administrator or operator can quickly power down a cluster or other infrastructure components with the click of a mouse.
- Synthetic WorkLoad (SWL) Rework - The SWL is an important acceptance testing component when new system architectures are deployed in LC. In recent years, we have expanded its scope to include testing new software releases on an ongoing basis. The current SWL suite is cumbersome, expert-friendly, and very difficult to maintain. With upcoming changes such as the release of TOSS-3, the SWL may not function at all in its current form. In CY15 LC will incorporate the tests in the SWL into a new, more friendly, and more adaptable testing structure. The Pavilion Test Harness (authored by Los Alamos) is the most promising of the technologies we are investigating for this new structure. Use of Pavilion would also affect a joint cooperative effort between LC, Los Alamos, and Sandia National Laboratories to produce a larger, shared suite of tests.
- HPSS Archive Hard Quota Implementation - Archival "soft" quotas have been used for many years to drastically reduce the amount of data sent to LC's archives. Unfortunately, some users continue to abuse their annual archive allocation requiring LC and Program staff to invest time and effort bringing them into compliance. In CY15 LC will develop and deploy hard archive quotas. After being automatically warned a configurable number of times, users heavily surpassing their archive quotas will be automatically blocked from creating new files within HPSS archives until they have freed space to a pre-established limit.

- Authentication Log Analysis and Monitoring - When a user attempts to authenticate to a system or service each discrete action is logged along with whether the action was successful or not. Monitoring and analyzing these logs is monotonous and time consuming. In CY15 LC will automate the processing of Center authentication logs.
- Automate Power Wall Alignment - Aligning projectors in LC's power walls is tedious and time consuming. During CY15 LC will begin using automated alignment software for its power wall projectors.
- Improve Infrastructure System Deployments - Deploying new infrastructure software releases is time consuming in large part because of the number of tests that need to be performed in order to insure proper operation and security are maintained. In CY15 LC will focus on the automation of checks on security services during upgrades and rollouts.
- Automated Archive Repack and Reclaim - Repacking data onto new or higher density existing archive tape media (repack and reclaim) is a constant and ongoing process. It is also very system administrator-intensive. During CY15 LC will undertake a project to develop and deploy an automated repack and reclaim utility.

## ***Delegation and Efficiency***

### *Issue:*

LC Hotline and Operations personnel regularly encounter issues and daily tasks in which they are required to, according to procedure, involve subject area experts even if the issue or task is simple and completely within the first responder's capability to handle.

### *Approach:*

In order to enhance efficiency LC will undertake the following tasks with the intention of empowering first responders to deal with common or well-defined tasks:

- Hotline Monitoring of Abusive Users - Hotline personnel are often on the receiving end of complaints about abusive users, particularly users abusing pdebug queues. Rather than waiting for users to complain, the Hotline will be trained and tasked with monitoring pdebug queues for abuse and with the notification of abusive users and their computer coordinators.
- Hotline NFS Quota Maintenance - The Hotline regularly fields NFS quota requests and is forced to engage system administrator personnel to adjust allocations. In CY15 the Hotline will be empowered to make NFS quota adjustments themselves.
- RSA One Time Password (OTP) Failure Issues - First responders in Operations, the Hotline, and 4-Help regularly deal with RSA OTP problems using the Hotpad utility. Often, there is not enough information provided by Hotpad to diagnose the problem being encountered. In CY15 LC will enhance Hotpad to allow for the diagnosis and repair of typical failure modes.
- Reworking the Degaussing Process - The process of degaussing both disk and tape media is extremely manpower intensive and process/paperwork-heavy. In CY15 Operations staff will review and rework the degaussing process and protocol for efficiency.

- Retiring CRYPTOCARD Tokens - Currently LC supports both CRYPTOCARD and RSA OTP solutions. This is time consuming and requires added infrastructure resources and staff expertise to support both solutions. During CY15 LC will investigate transitioning to the use of RSA OTP tokens only, enabling retirement of CRYPTOCARD services.
- Simplifying Atlassian Authenticators - The Atlassian tool suite currently employs three different custom authenticators, all of which need to be understood and maintained. In CY15 LC will retire these authenticators and replace them with Atlassian Crowd identity management authentication.
- Reduce Frequency of Infrastructure TOSS updates - The number and complexity of LC infrastructure systems makes it extremely difficult to keep pace with TOSS operating system releases installed on LC production clusters. In CY15 LC will study reducing the frequency of TOSS updates to infrastructure machines to once every six months and, if found feasible, LC will implement this policy.
- External Cloud-Based Testing of Software - LC software developers have long been involved in developing open source software. Two such on-going efforts are ZFS on Linux (ZoL) and Flux. Both of these efforts are hosted on GitHub with ZoL, in particular, having a large number of external collaborators. Testing and reviewing of patches submitted by such external collaborators can be improved by making use of external cloud-based VM environments. During CY15 LC will investigate the practicality and cost tradeoffs of using external cloud-based systems for testing of ZoL and Flux open source software stacks. If it is determined that LC should move forward, then a six-month pilot project will be undertaken. Afterwards, cost and efficiency tradeoffs of cloud-based options will be reviewed.
- Investigation of IDM Replacement - LC's existing identity management (IDM) infrastructure is largely proprietary, highly complex, and is no longer sold or supported. In CY15 LC will investigate connecting a collection of software products to replace the existing IDM workflow, approvals, and provisioning processes.
- Automating Deletion of Empty Storage Accounts - When a user's account is retired, Hotline Account Specialists coordinate with Data Storage Group (DSG) personnel to delete (often empty) archive accounts. In CY15 the deletion of empty storage accounts will be automated, minimizing the need to coordinate between Center staff, and streamlining the workflow of de-provisioning archive accounts.
- Offloading TotalView Tasks - With the retirement of the person responsible for all TotalView debugger issues, the Development Environment Group (DEG) was forced to absorb TotalView responsibilities within remaining staff. In CY15 LC will establish a contract with Rogue Wave Software to handle many of these tasks, freeing up DEG staff to work on high priority support issues.
- Operations Training Cluster - Given proper training and hands-on experience, Operations personnel would be able take on an even greater workload of tasks typically performed by system administrators. In CY15 LC will deploy a small training cluster and use it to train operators in system administration tasks including the deployment and decommissioning of

clusters. This will lead to the delegation of additional system administration tasks to Operations as appropriate.

## **Next Generation Compute and Data Environments**

CTS-1, ATS-2 and pre-exascale machine horizons, and the rise of big data and data analytics disciplines, necessitate that LC invest in large-scale preparation efforts. These efforts include architectural research, targeted software development, and production data analytics system administration and support. Supporting future machines and their architectures will require scaling efforts be performed on most HPC infrastructure components. Prospective new mission support areas, including data science technologies, require LC to rapidly grow its expertise in non-traditional HPC spaces.

### ***CTS-1 Preparation***

#### *Issue:*

Following the procurement precedents set by the preceding Tri-Laboratory Linux Capacity Cluster procurements (TLCC-1 and TLCC-2) the Tri-Laboratories will go out to bid for the first Commodity Technology Systems (CTS-1). While the machines will not arrive until CY16, much effort must be invested to complete selection, procurement, and preparation/planning phases in CY15.

#### *Approach:*

In preparation for the CTS-1 procurement, the CTS-1 Statement Of Work (SOW) and Request For Proposal (RFP) will be completed and all necessary approvals will be obtained. The CTS-1 competition will be completed and LC personnel along with ASC, SNL and LANL personnel will select a winning solution and will establish a contract for the resulting machines.

In preparation for the CY16 arrival of CTS-1 hardware at LLNL, LC will develop and release the TOSS-3 operating system release which will be a major release based on Red Hat Enterprise Linux version 7 (RHEL 7). Network planning and preparation will be undertaken in advance of the activation of B654 - the building which will house CTS-1 equipment. LC file system staff will investigate and determine the new hardware System Scalable Unit (SSU) to be deployed in file systems procured in support of CTS-1 systems. HPSS developers and administrators will test an early release of HPSS 7.5 and its metadata partitioning features which are planned for production support in the CTS-1 timeframe.

### ***Sierra Preparation***

#### *Issue:*

The Sierra ATS-2 machine was selected in CY14 and the machine is scheduled to arrive late in CY17. Technology demonstration platforms will arrive at contract-defined points over the intervening two years. While the initial Sierra Non-Recurring Engineering (NRE) and build contracts were established in CY14, the amount of work and collaboration with IBM required to enable platform success is daunting and will continue to involve every discipline within LC throughout CY15.

#### *Approach:*

A multi-year NRE contract was established in CY15 in order to advance hardware and software technologies at a pace that will support timely delivery of the Sierra machine. This contract, and its

milestones, describes the research and development needed to accelerate the state-of-the art in software and hardware solutions bid in the build contract. This very large contract has required, and will continue to require, very significant weekly involvement from staff across LC and the Advanced Technology Office (ATO). The importance and volume of effort required to collaborate with the Sierra vendor partner and to influence the research and architectures involved in the NRE contract should not be underestimated.

The Sierra machine will include GPUs along with traditional IBM POWER processors. The level of skill and amount of effort to convert existing codes to efficiently utilize Sierra GPUs to their fullest potential will be extreme. In order to jump-start ASC and other code preparation for Sierra, LC and IBM have established a Center Of Excellence (COE). In CY15, the COE will be extremely active, and will both educate and require the focus of LC staff. In support of this effort, LC will begin GP-GPU training of its employees and work to provide platforms and environments enabling initial testing and education surrounding the integration of GPUs into applications.

## ***Pre-Exascale Preparation***

### *Issue:*

It is unclear which HPC architectures will be dominant or even viable in the extended pre-exascale and exascale timeframes. HPC vendors themselves have yet to settle on their targeted architectures for the 2020-2025 time period. Understanding and narrowing the solution space early, and identifying key technological hurdles will be necessary so that application and infrastructure teams can adequately prepare for the arrival and efficient use of future platforms.

### *Approach:*

Within LC, two long-horizon preparation efforts will be undertaken. The first is the continuation of the Flux development effort. Flux is an extensible, next-generation job and resource management framework that will target management and scheduling of the entire computer center as a single pool of resources. Flux development and prototyping will continue throughout CY15. The second long-horizon effort will be the investigation of file system alternatives in the post-Sierra timeframe. LC staff will survey present and prospective file system solutions in order to inform long-term Center architecture and software planning efforts.

ICCD's Advanced Technology Office (ATO) will team closely with LC discipline experts and customers in investigations and vendor engagements surrounding next generation platform technologies. Investigations will be performed on POWER, ARM, Intel Xeon Phi, GPU and Intel x86 microarchitectures. LC will continue active involvement and participation in all DOE FastForward2 and DesignForward2 projects and will work to establish an additional FastForward2 project in the area of I/O and storage. LC and ATO personnel will be heavily involved in the Storage System and I/O (SSIO) research efforts driven by the DOE Office of Science. Together these efforts will focus on accelerating the deployment of pre-exascale and exascale machines and their associated software.

## ***Big Data***

### *Issue:*

There has been a global explosion in the use of data science techniques to analyze large amounts of typically unstructured data ("Big Data" gathered from social networks, sensor networks, etc.) into valuable and actionable data products. Over the last two years LC has promoted focused exploratory efforts surrounding Big Data with an eye toward integration with our existing HPC

architectures. LC also established the Catalyst machine environment in part to investigate new approaches to big data analytics. Despite these efforts, LC has yet to generate a suite of big data services for users.

*Approach:*

In CY15 LC will continue to prototype and investigate ways of leveraging and growing our HPC and big data expertise for use in data science frameworks, especially in the area of alternative underpinnings to Hadoop. LC will then continue exploring opportunities to improve Hadoop performance with HPC technologies including Lustre, InfiniBand, and NVRAM and will leverage the Catalyst machine environment and Magpie projects in these efforts. LC will also provide ongoing support for clusters, storage and big data platforms for external programs located within LC buildings. Initial support will be provided for Global Security hardware and equipment in support of SNSI-level efforts.

The first substantive customers of LC big data expertise may indeed be the LC itself. LC clusters and infrastructure systems generate huge volumes of log information. The addition of LDMS to the LC's logging and monitoring suite introduces the distinct likelihood of generating far more log information than the LC can parse into actionable intelligence. In CY15 LC will develop a prototype hardware and software architecture that can continuously collect and analyze system data produced by LDMS along with performance measurements from Weapons and Complex Integration (WCI) proxy applications. Continuous monitoring will provide both staff and customers with increased insight into system and application performance in our dynamic and variable production environment. This insight in turn will facilitate diagnosis of problems such as Lustre performance, network issues, and application performance issues for WCI codes. It will also aid performance data collection and analysis efforts on the WCI Next Generation Code Project.